GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:47:19 ; Search time 210.42 Seconds
(without alignments)
9.750 Million cell updates/sec

Title: US-09-331-631A-7_COPY_81_140
Perfect score: 342
Sequence: 1 LQRQYQQCQGRCQEQQQGQR.....HENYHNHKKNRSEEEEGQQR 60

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	342	100.0	525	19	W62831	Theobroma cacao an
2	342	100.0	566	13	R20181	Sequence encoded b
3	133	38.9	590	19	W62832	Gossypium hirsutum
4	130.5	38.2	666	19	W62829	Macadamia integrif
5	127.5	37.3	625	19	W62830	Macadamia integrif
6	126.5	37.0	666	19	W62828	Macadamia integrif
7	92	26.9	2074	21	Y54319	Amino acid sequenc
8	87	25.4	482	20	Y07067	Renal cancer assoc
9	86.5	25.3	395	17	W03474	Mouse SRY-related
10	86	25.1	2023	21	Y54320	Amino acid sequenc
11	85.5	25.0	611	20	Y29039	T. gondii immunoge
12	84.5	24.7	86	20	W95073	GST-HD fusion prot
13	84.5	24.7	86	20	W95078	GST-HD fusion prot
14	84.5	24.7	94	20	W95075	GST-HD fusion prot
15	84.5	24.7	94	20	W95080	GST-HD fusion prot
16	83.5	24.4	919	10	P93109	Human androgen rec
17	83.5	24.4	919	18	W14783	Androgen receptor.
18	83.5	24.4	919	21	Y78914	Human androgen rec
19	82	24.0	28	19	W62841	Stenocarpus sinuat
20	82	24.0	910	20	Y22191	Mouse brain CNG-1
21	81	23.7	371	20	W73369	Epitope tagged TBP
22	81	23.7	593	19	W62835	Zea mays antimicro
23	81	23.7	1162	21	Y58500	HHV8 ORF 73 protei
24	80	23.4	562	16	R70491	Leucocytozoan prot
25	80	23.4	1898	20	Y30795	A human trichohyal
26	78.5	23.0	360	17	W03627	Human follicle sti
27	76	22.2	637	19	W62837	Hordeum vulgare an
28	76	22.2	1326	20	Y55933	Human ZC3 protein.
29	75	21.9	1178	18	W30763	Mannose-1-phosphat
30	74.5	21.8	326	20	Y20109	B. burgdorferi ant
31	74.5	21.8	347	20	Y20108	B. burgdorferi ant
32	74.5	21.8	616	20	Y32013	Drosophila melanog
33	74	21.6	108	20	W95071	Amino acid sequenc
34	74	21.6	108	20	W95076	Amino acid sequenc
35	73.5	21.5	542	14	R38746	fySSRP. Saccharom
36	73.5	21.5	542	19	W39214	S. cerevisiae SSRP
37	73.5	21.5	1299	21	Y58633	Protein regulating
38	72.5	21.2	409	20	W90342	G. max truncated S
39	72.5	21.2	489	20	W90341	G. max SBP2 protei
40	72.5	21.2	1297	20	Y55932	Human ZC2 protein.
41	72.5	21.2	1447	20	W81029	Murine pCIP protei
42	72	21.1	154	20	Y33504	Human unliganded a
43	72	21.1	444	20	W90340	G. max truncated S
44	72	21.1	521	19	W74802	Human secreted pro
45	72	21.1	524	20	W90339	G. max SBP1 protei

ALIGNMENTS

RESULT 1
W62831
ID W62831 standard; Protein; 525 AA.
XX
AC W62831;
XX
DT 27-OCT-1998 (first entry)
XX
DE Theobroma cacao antimicrobial protein.
XX
KW antimicrobial protein; infestation; control.
XX
OS Theobroma cacao.
XX
PN WO9827805-A1.
XX
PD 02-JUL-1998.
XX
PF 22-DEC-1997; 97WO-AU00874.
XX
PR 20-DEC-1996; 96AU-0004275.
XX
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX
DR WPI; 1998-377279/32.
XX
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
XX
PS Claim 1; Page 47-49; 96pp; English.
XX
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian

```
CC   animals.
XX
SQ   Sequence   525 AA;


  Query Match            100.0%;  Score 342;  DB 19;  Length 525;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-29;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LQRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 lqrqyqqcqgrcqeqqqgqreqqqcqrkcweqykeqergehenyhnhkknrseeeegqqr 140


RESULT   2
R20181
ID   R20181 standard; Protein; 566 AA.
XX
AC   R20181;
XX
DT   16-APR-1992  (first entry)
XX
DE   Sequence encoded by 67 kD T. cacao protein cDNA.
XX
KW   Cocoa; flavour; vicilin; seed storage protein.
XX
OS   Theobroma cacao.
XX
PN   WO9119801-A.
XX
PD   26-DEC-1991.
XX
PF   07-JUN-1991;   91WO-GB00914.
XX
PR   11-JUN-1990;   90GB-0013016.
XX
PA   (MRSC ) MARS UK LTD.
XX
PI   Spencer ME,  Hodge R,  Deakin EA,  Ashton S;
XX
DR   WPI; 1992-024418/03.
DR   N-PSDB; Q20377.
XX
PT   Recombinant cocoa proteins - are responsible for flavour in cocoa
PT   beans and produced in large quantities using yeast and bacterial
PT   expression vectors
XX
PS   Claim 4; Fig 2; 59pp; English.
XX
CC   The inventors claim a 67 kD and 31 kD T. cacao protein, and
CC   fragments, and encoding DNAs.  The 47 kD and 31 kD proteins are
CC   derived from the 67 kD precursor.  T. cacao protein cDNA was
CC   detected in a cDNA library prepared from immature cocoa beans RNA
CC   using a probe based on the AA sequence of a CNBr peptide common to
CC   the 47 kD and 31 kD polypeptides.  Homology searches revealed close
CC   homologies between the 67 kD polypeptide and the vicilins, which are
CC   seed storage proteins.
XX
SQ   Sequence   566 AA;


  Query Match            100.0%;  Score 342;  DB 13;  Length 566;
  Best Local Similarity  100.0%;  Pred. No. 2e-29;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LQRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 lqrqyqqcqgrcqeqqqgqreqqqcqrkcweqykeqergehenyhnhkknrseeeegqqr 140


RESULT   3
W62832
ID   W62832 standard; Protein; 590 AA.
XX
AC   W62832;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Gossypium hirsutum antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Gossypium hirsutum.
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 49-51; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   590 AA;


  Query Match            38.9%;  Score 133;  DB 19;  Length 590;
  Best Local Similarity  35.8%;  Pred. No. 7.3e-07;
  Matches   29;  Conservative   12;  Mismatches   18;  Indels   22;  Gaps    3;

Qy      2 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQ--------------ERGEHENYHN 46
          |||:|:||  | :|:|   ::||| |:| |:|:|               | || |  ||
Db    121 qrqfqecqqhchqqeqrpekkqqcvrecrekyqenpwrgereeeaeeeeteegeqeqshn 180

Qy     47 ----HKK---NRSEEEEGQQR 60
              |::   :|  || |  |
Db    181 pfhfhrrsfqsrfreehgnfr 201


RESULT   4
W62829
ID   W62829 standard; Protein; 666 AA.
XX
AC   W62829;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Macadamia integrifolia antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Macadamia integrifolia.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "signal peptide"
FT   Protein         29..666
FT                   /note= "mature protein"
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
```

```
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
DR   N-PSDB; V42311.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 39-41; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   666 AA;


  Query Match             38.2%;  Score 130.5;  DB 19;  Length 666;
  Best Local Similarity   34.3%;  Pred. No. 1.5e-06;
  Matches   24;  Conservative   12;  Mismatches   23;  Indels   11;  Gaps    1;

Qy       2 QRQYQQCQGRCQEQQQGQREQQQCQRKC----------WEQYKEQERGEHENYHNHKKN 50
           |:||:||| |||  :   |  | ||::|           ::|:||:| : | |    |
Db     123 qqqyeqcqercqrheteprhmqtcqqrcerryekekrkqqkryeeqqredeekyeermke 182

Qy      51 RSEEEEGQQR 60
             : : |||
Db     183 ednkrdpqqr 192


RESULT   5
W62830
ID   W62830 standard; Protein; 625 AA.
XX
AC   W62830;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Macadamia integrifolia antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Macadamia integrifolia.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "signal peptide"
FT   Protein         29..666
FT                   /note= "mature protein"
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
DR   N-PSDB; V42316.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 43-45; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   625 AA;


  Query Match             37.3%;  Score 127.5;  DB 19;  Length 625;
  Best Local Similarity   34.3%;  Pred. No. 3e-06;
  Matches   24;  Conservative   13;  Mismatches   22;  Indels   11;  Gaps    1;

Qy       2 QRQYQQCQGRCQEQQQGQREQQQCQRKC----------WEQYKEQERGEHENYHNHKKN 50
           |:||:||| |||  :   |  | ||::|           ::|:||:| : | |    |
Db      82 qqqyeqcqkrcqrreteprhmqicqqrcerryekekrkqqkryeeqqredeekyeermke 141

Qy      51 RSEEEEGQQR 60
             : : |||
Db     142 gdnkrdpqqr 151


RESULT   6
W62828
ID   W62828 standard; Protein; 666 AA.
XX
AC   W62828;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Macadamia integrifolia antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Macadamia integrifolia.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "signal peptide"
FT   Protein         29..666
FT                   /note= "mature protein"
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
DR   N-PSDB; V42310.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 34-36; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   666 AA;


  Query Match             37.0%;  Score 126.5;  DB 19;  Length 666;
  Best Local Similarity   32.9%;  Pred. No. 4.2e-06;
```

```
  Matches   23;  Conservative   13;  Mismatches   23;  Indels   11;  Gaps    1;

Qy       2 QRQYQQCQGRCQEQQQGQREQQQCQRKC----------WEQYKEQERGEHENYHNHKKN 50
           |:||:|||  || ::   |  | ||::|           ::|:||:| : | |    |
Db     123 qqqyeqcqkhcqrreteprhmqtcqqrcerryekekrkqqkryeeqqredeekyeermke 182

Qy      51 RSEEEEGQQR 60
             : : |||
Db     183 ednkrdpqqr 192


RESULT   7
Y54319
ID   Y54319 standard; Protein; 2074 AA.
XX
AC   Y54319;
XX
DT   06-APR-2000  (first entry)
XX
DE   Amino acid sequence of a murine PCTG4 protein.
XX
KW   Human; PCTG4 region; X chromosome; q13 region; polymorphism;
KW   mental retardation; autism; depression; bipolar affective disorder;
KW   hypothyroidism; OPA gene; neuropsychiatric disorder.
XX
OS   Mus sp.
XX
PN   WO9955915-A2.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09365.
XX
PR   29-APR-1998;   98US-0083465.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Philibert RA,  Ginns EI;
XX
DR   WPI; 2000-126357/11.
XX
PT   Identification of polymorphisms in the PCTG4 region of Xq13 for
PT   diagnosing mental retardation or autism -
XX
PS   Example 7; Page 81-84; 100pp; English.
XX
CC   The present sequence represents a murine PCTG4 protein. Polymorphisms
CC   in the human PCTG4 region of chromosome Xq13 are associated with
CC   mental retardation, autism, depression, bipolar affective disorder or
CC   hypothyroidism. One 12 bp insertion polymorphism occurs within the
CC   coding region of the human OPA gene, and introduces a 4 amino acid
CC   insertion in a putative OPA domain. This domain has been shown to be
CC   involved in tissue specific expression. Another polymorphism consists
CC   of a pentanucleotide repeat approximately 7 kb upstream of the 12 bp
CC   polymorphism. Another polymorphisms consists of a dinucleotide repeat
CC   approximately 4.5 kb downstream of the 12 bp polymorphism. The
CC   specification describes a method for screening for polymorphisms in a
CC   PCTG4 nucleic acid sequence obtained from a subject. The PCTG4 related
CC   sequences within the q13 region of the X chromosome have polymorphisms
CC   associated with neuropsychiatric disorders.  The methods can be used to
CC   screen for the presence of a heritably linked form of mental retardation,
CC   autism, depression, bipolar affective disorder or hypothyroidism.
XX
SQ   Sequence   2074 AA;


  Query Match             26.9%;  Score 92;  DB 21;  Length 2074;
  Best Local Similarity   30.5%;  Pred. No. 0.073;
  Matches   18;  Conservative   21;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       2 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 60
           |:| || | : |:||| |::||| |::  :||  ::: : :     :: : :::: ||:

Db    1942 qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqyhirqqqqqqqmlrqqqqqqqqqqqqq 2000


RESULT   8
Y07067
ID   Y07067 standard; Protein; 482 AA.
XX
AC   Y07067;
XX
DT   02-JUL-1999  (first entry)
XX
DE   Renal cancer associated antigen precursor sequence.
XX
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer.
XX
OS   Homo sapiens.
XX
PN   WO9904265-A2.
XX
PD   28-JAN-1999.
XX
PF   15-JUL-1998;   98WO-US14679.
XX
PR   22-JUN-1998;   98US-0102322.
PR   17-JUL-1997;   97US-0896164.
PR   10-OCT-1997;   97US-0061599.
PR   10-OCT-1997;   97US-0061765.
PR   10-OCT-1997;   97US-0948705.
PR   11-OCT-1997;   97GB-0021697.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y,  Gout I,  Gure A,  O'Hare M,  Obata Y,  Old LJ;
PI   Pfreundschuh M,  Sahin U,  Scanlan MJ,  Stockert E;
PI   Tureci O;
XX
DR   WPI; 1999-132448/11.
XX
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
XX
PS   Disclosure; Page 467-468; 787pp; English.
XX
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
XX
SQ   Sequence   482 AA;


  Query Match             25.4%;  Score 87;  DB 20;  Length 482;
  Best Local Similarity   25.4%;  Pred. No. 0.054;
  Matches   15;  Conservative   22;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       2 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 60
           |:  :: : : ::::  :: |:: :||  ||::||:| | |     :: : |:||  |:
Db      80 qkaqeefrikkekeeaakkrqeeqerklkeqweeqqrkereeeeqkrqekkekeealqk. 138
```

```
RESULT   9
W03474
ID   W03474 standard; Protein; 395 AA.
XX
AC   W03474;
XX
DT   23-OCT-1996  (first entry)
XX
DE   Mouse SRY-related protein.
XX
KW   Mouse; SRY; primer; PCR; polymerase chain reaction; amplification; probe;
KW   HMG box; human; bovine; sex; animal; birth.
XX
OS   Mus musculus.
XX
PN   JP08154685-A.
XX
PD   18-JUN-1996.
XX
PF   30-NOV-1994;   94JP-0319525.
XX
PR   30-NOV-1994;   94JP-0319525.
XX
PA   (KACH-) KACHIKU JUSEIRAN ISHOKU GIKUTSU KENKYUKU.
XX
DR   WPI; 1996-336575/34.
DR   N-PSDB; T33007.
XX
PT   Bovine and mouse Sry-related DNA - useful for detecting e.g. the sex
PT   of unborn animals
XX
PS   Claim 1; Page 10-14; 21pp; Japanese.
XX
CC   This is the amino acid sequence of a mouse SRY-related protein.  The gene
CC   was isolated from a mouse genomic library using a cDNA fragment amplified
CC   by primers T33009-10 as a probe.  The screen isolated 4 EcoRI fragments
CC   of 2.3, 2.8, 3.5 and 1.5 kb covering the gene. Sequence analysis revealed
CC   a 240 bp HMG box sequence between bases 7154-7393.  Similarity with the
CC   human SRY HMG box sequence resulted in primers being generated to amplify
CC   the human SRY HMG box sequence for use as a probe to isolate the bovine
CC   SRY-related gene (T33008).  The mouse and bovine genes are useful for
CC   determining the sex of an animal prior to birth.
XX
SQ   Sequence   395 AA;


  Query Match             25.3%;  Score 86.5;  DB 17;  Length 395;
  Best Local Similarity   27.0%;  Pred. No. 0.049;
  Matches   17;  Conservative   21;  Mismatches   20;  Indels    5;  Gaps    1;

Qy       3 RQYQQCQGRCQEQQQGQREQQQC-----QRKCWEQYKEQERGEHENYHNHKKNRSEEEEG 57
           :| ||     |:||| |::|||      ::: :  : :|::  |:::|:|:: :  :
Db     160 qqqqqfydhhqqqqqqqqqqqfhdhhqekqefhdhhqqqqqfhdhhhhhqeqqfhdhhq 219

Qy      58 QQR 60
           ||:
Db     220 qqq 222


RESULT  10
Y54320
ID   Y54320 standard; Protein; 2023 AA.
XX
AC   Y54320;
XX
DT   06-APR-2000  (first entry)
XX
DE   Amino acid sequence of a human PCTG4 protein.
XX
KW   Human; PCTG4 region; X chromosome; q13 region; polymorphism;
KW   mental retardation; autism; depression; bipolar affective disorder;
KW   hypothyroidism; OPA gene; neuropsychiatric disorder.
XX
OS   Homo sapiens.
XX
PN   WO9955915-A2.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09365.
XX
PR   29-APR-1998;   98US-0083465.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Philibert RA,  Ginns EI;
XX
DR   WPI; 2000-126357/11.
XX
PT   Identification of polymorphisms in the PCTG4 region of Xq13 for
PT   diagnosing mental retardation or autism -
XX
PS   Example 7; Page 81-84; 100pp; English.
XX
CC   The present sequence represents a human PCTG4 protein. Polymorphisms
CC   in the human PCTG4 region of chromosome Xq13 are associated with
CC   mental retardation, autism, depression, bipolar affective disorder or
CC   hypothyroidism. One 12 bp insertion polymorphism occurs within the
CC   coding region of the human OPA gene, and introduces a 4 amino acid
CC   insertion in a putative OPA domain. This domain has been shown to be
CC   involved in tissue specific expression. Another polymorphism consists
CC   of a pentanucleotide repeat approximately 7 kb upstream of the 12 bp
CC   polymorphism. Another polymorphisms consists of a dinucleotide repeat
CC   approximately 4.5 kb downstream of the 12 bp polymorphism. The
CC   specification describes a method for screening for polymorphisms in a
CC   PCTG4 nucleic acid sequence obtained from a subject. The PCTG4 related
CC   sequences within the q13 region of the X chromosome have polymorphisms
CC   associated with neuropsychiatric disorders.  The methods can be used to
CC   screen for the presence of a heritably linked form of mental retardation,
CC   autism, depression, bipolar affective disorder or hypothyroidism.
XX
SQ   Sequence   2023 AA;


  Query Match             25.1%;  Score 86;  DB 21;  Length 2023;
  Best Local Similarity   30.5%;  Pred. No. 0.32;
  Matches   18;  Conservative   18;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 60
           |:| || | : |:||| |::||| |    :| ::|   : :     :: : :::: ||:
Db    1899 qqqqqqqqqqqqqqqqqqqqqqqqqyhirqqqqqqilrqqqqqqqqqqqqqqqqqqqq 1957


RESULT  11
Y29039
ID   Y29039 standard; Protein; 611 AA.
XX
AC   Y29039;
XX
DT   24-SEP-1999  (first entry)
XX
DE   T. gondii immunogenic protein.
XX
KW   Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;
KW   T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;
KW   Toxoplasma oocyst.
XX
OS   Toxoplasma gondii.
XX
PN   WO9932633-A1.
XX
PD   01-JUL-1999.
XX
PF   18-DEC-1998;   98WO-US27137.
XX
```

PR 19-DEC-1997; 97US-0994825.
XX
PA (HESK-) HESKA CORP.
XX
PI Lutz SB, Milhausen MJ, Ng RK;
XX
DR WPI; 1999-418930/35.
DR N-PSDB; X91242.
XX
PT New isolated Toxoplasma gondii nucleic acids used, e.g. to treat
PT infection caused by this microorganism
XX
PS Claim 29; Page 227-229; 381pp; English.
XX
CC The invention provides isolated Toxoplasma gondii nucleic acids that
CC encode immunogenic polypeptides. The T. gondii nucleic acid molecules,
CC immunogenic proteins and antibodies to the proteins can be used to
CC inhibit T. gondii oocyst shedding in a cat due to infection with
CC T. gondii. They can be used for preventing T. gondii infection and for
CC preventing the spread of T. gondii infection. They can also be used for
CC detecting T. gondii infection. The detection method can be used to detect
CC parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts
CC such as Cryptosporidium oocysts and Toxoplasma oocysts.
XX
SQ Sequence 611 AA;

Query Match 25.0%; Score 85.5; DB 20; Length 611;
Best Local Similarity 34.5%; Pred. No. 0.1;
Matches 19; Conservative 14; Mismatches 17; Indels 5; Gaps 1;

Qy 11 RCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNH-----KKNRSEEEEGQQR 60
Db 447 rveeekarqreeeererrrveeekererqeeererrrveeekererqeeeererr 501

RESULT 12
W95073
ID W95073 standard; Protein; 86 AA.
XX
AC W95073;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELP.
XX
KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX
PN WO9906838-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP04810.
XX
PR 01-AUG-1997; 97EP-0113320.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153955/13.
XX
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for
PT diagnosis, particularly of diseases associated with polyglutamine
PT expansion
XX
PS Disclosure; Fig 8; 56pp; English.
XX
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences W95072-75 represent GST-HD fusion proteins.
XX
SQ Sequence 86 AA;

Query Match 24.7%; Score 84.5; DB 20; Length 86;
Best Local Similarity 28.1%; Pred. No. 0.016;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 1;

Qy 3 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 59
Db 22 ksfqqqq---qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 75

RESULT 13
W95078
ID W95078 standard; Protein; 86 AA.
XX
AC W95078;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELP.
XX
KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW GST-HD; HD.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX
PN WO9906545-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP04811.
XX
PR 01-AUG-1997; 97EP-0113306.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153775/13.
XX
PT Composition containing fusion protein that includes amyloidogenic
PT peptide - able to self-assemble into fibrils or aggregates, used to
PT detect and monitor neuronal diseases, and also to screen for
PT therapeutic inhibitors
XX
PS Disclosure; Fig 8; 62pp; English.
XX
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences W95077-80 represent GST-HD fusion
CC proteins.
XX
SQ Sequence 86 AA;

Query Match 24.7%; Score 84.5; DB 20; Length 86;
Best Local Similarity 28.1%; Pred. No. 0.016;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 1;

```
Qy      3 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 59
          : :|| |   |:||| |::||| |::  :| ::|:: : :     :: : :::: ||
Db     22 ksfqqqq---qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 75
```

RESULT 14
W95075
ID W95075 standard; Protein; 94 AA.
XX
AC W95075;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELPBio.
XX
KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX
PN WO9906838-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP04810.
XX
PR 01-AUG-1997; 97EP-0113320.

XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153955/13.
XX
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for
PT diagnosis, particularly of diseases associated with polyglutamine
PT expansion
XX
PS Disclosure; Fig 8; 56pp; English.
XX
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences W95072-75 represent GST-HD fusion proteins.
XX
SQ Sequence 94 AA;

Query Match 24.7%; Score 84.5; DB 20; Length 94;
Best Local Similarity 28.1%; Pred. No. 0.018;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 1;

```
Qy      3 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 59
          : :|| |   |:||| |::||| |::  :| ::|:: : :     :: : :::: ||
Db     22 ksfqqqq---qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 75
```

RESULT 15
W95080
ID W95080 standard; Protein; 94 AA.
XX
AC W95080;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELPBio.
XX
KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW GST-HD; HD.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX
PN WO9906545-A2.
XX
PD 11-FEB-1999.

XX
PF 31-JUL-1998; 98WO-EP04811.
XX
PR 01-AUG-1997; 97EP-0113306.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153775/13.
XX
PT Composition containing fusion protein that includes amyloidogenic
PT peptide - able to self-assemble into fibrils or aggregates, used to
PT detect and monitor neuronal diseases, and also to screen for
PT therapeutic inhibitors
XX
PS Disclosure; Fig 8; 62pp; English.
XX
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences W95077-80 represent GST-HD fusion
CC proteins.
XX
SQ Sequence 94 AA;

Query Match 24.7%; Score 84.5; DB 20; Length 94;
Best Local Similarity 28.1%; Pred. No. 0.018;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 1;

```
Qy       3 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 59
           : :|| |   |:||| |::||| |::  :| ::|:: : :     :: : :::: ||
Db      22 ksfqqqq---qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 75
```

Search completed: March 1, 2001, 15:47:20
Job time: 245 sec